Q1-1: Are these statements true or false?
(A) If we have multiple optimal solutions on a given training set, those solutions will also have the same test loss.
(B) If a hyperplane only changes its bias term by 1, then the distance from some point x to the hyperplane will not change.

1. True, True
2. True, False
3. False, True
4. False, False

Q1-1: Are these statements true or false?
(A) If we have multiple optimal solutions on a given training set, those solutions will also have the same test loss.
(B) If a hyperplane only changes its bias term by 1, then the distance from some point x to the hyperplane will not change.

1. True, True
2. True, False
3. False, True
4. False, False ←

(A) Multiple optimal solutions on the training usually have different test loss. Please refer to the example given in the lecture.
(B) Recall that the distance is given by $\frac{|f_{w,b}(x)|}{\|w\|}$. If only the bias term is changed, then $|f_{w,b}(x)|$ will change while $\| w \|$ remains same. So the distance will also be changed.

Q1-2: Please calculate the distance from $x = (1, 2, 3)$ to the hyperplane with $w = (1, 2, 2)$ and $b = -1$. Recall that the distance is given by $\frac{|f_{w,b}(x)|}{\|w\|}$.

1. $10/9$
2. $4/3$
3. $10/3$
4. 4

Q1-2: Please calculate the distance from $x = (1, 2, 3)$ to the hyperplane with $w = (1, 2, 2)$ and $b = -1$. Recall that the distance is given by $\frac{|f_{w,b}(x)|}{\|w\|}$.

1. $10/9$
2. $4/3$
3. $10/3$
4. 4

Recall that the distance is given by $\frac{|f_{w,b}(x)|}{\|w\|}$. Then we have $\mathrm{d} = \frac{|f_{w,b}(x)|}{\|w\|} = \frac{|1*1+2*2+2*3-1|}{\sqrt{1^2+2^2+2^2}} = \frac{10}{3}$.

Q2-1: Are these statements true or false?

(A) Define the margin to be $\gamma = \min_i \frac{y_i f_{w,b}(x_i)}{\|w\|}$, if $f_{w,b}(\mathrm{x})$ predicts correctly on some x_i and incorrectly on others, then the margin will be positive.

(B) If the training set can be correctly separated, then $\max_{w,b} \gamma$ can still be negative.

1. True, True
2. True, False
3. False, True
4. False, False

Q2-1: Are these statements true or false?

(A) Define the margin to be $\gamma = \min_i \frac{y_i f_{w,b}(x_i)}{\|w\|}$, if $f_{w,b}(\mathrm{x})$ predicts correctly on some x_i and incorrectly on others, then the margin will be positive.

(B) If the training set can be correctly separated, then $\max_{w,b} \gamma$ can still be negative.

1. True, True
2. True, False
3. False, True
4. False, False

(A) In this case, $\frac{y_i f_{w,b}(x_i)}{\|w\|}$ would be negative on those x_i with incorrect predictions. So take min on all training data, we will get the margin negative.

(B) In this case, there exists at least one w and b such that all instances are correctly classified, so the corresponding margin is non-negative.

Q2-2: Are these statements true or false?
(A) Assume that a hyperplane has parameters (w, b) and the margin of the training set is γ. If we change (w, b) to $(0.5w, 0.5b)$, then the margin will become 2γ.
(B) Consider a fixed scale such that $y_i f_{w,b}(x_i) \geq 2$ and the equality holds for at least one point, then the margin $\gamma = \frac{1}{\|w\|}$, and maximizing the margin on the training set is equivalent to minimizing $\| w \|$.

1. True, True
2. True, False
3. False, True
4. False, False

Q2-2: Are these statements true or false?

(A) Assume that a hyperplane has parameters (w, b) and the margin of the training set is γ. If we change (w, b) to $(0.5w, 0.5b)$, then the margin will become 2γ.

(B) Consider a fixed scale such that $y_i f_{w,b}(x_i) \geq 2$ and the equality holds for at least one point, then the margin $\gamma = \frac{1}{\|w\|}$, and maximizing the margin on the training set is equivalent to minimizing $\| w \|$.

1. True, True
2. True, False
3. False, True
4. False, False

(A) It is showed in the lecture that the margin is invariant to scaling of (w, b).

(B) Here the margin should be $\frac{2}{\|w\|}$. Also, maximizing the margin on the training set is equivalent to minimizing $\| w \|$ subject to the constraints that $y_i f_{w,b}(x_i) \geq 2$.

Q3-1: Are these statements true or false?
(A) The solution of SVM will always change if we remove some instances from the training set.
(B) If we can only access the labels and the inner products of instances $\{x_i{}^T x_j\}_{i,j}$, we can NOT solve the learning problem in SVM.

1. True, True
2. True, False
3. False, True
4. False, False

Q3-1: Are these statements true or false?
(A) The solution of SVM will always change if we remove some instances from the training set.
(B) If we can only access the labels and the inner products of instances $\{x_i{}^T x_j\}_{i,j}$, we can NOT solve the learning problem in SVM.

1. True, True
2. True, False
3. False, True
4. False, False

(A) As is shown in the lecture, if we remove those instances with $\alpha_i = 0$, it will not influence the SVM result.
(B) We can see that the dual problem only depends on y_i and the inner products of training instances. So we can also solve the SVM problem in this case.

Q3-2: Please solve the learning problem of SVM to get w for the following training data ($x \in \mathbb{R}^2$, $\mathrm{y} \in \mathbb{R}$). Hint: what are the support vectors?

$x^{(1)}$	$x^{(2)}$	y
3	3	1
4	4	1
1	1	-1

1. $(0, 1)$
2. $(1,0)$
3. $(\frac{1}{2}, \frac{1}{2})$
4. $(\frac{6}{13}, \frac{4}{13})$

Q3-2: Please solve the learning problem of SVM to get w for the following training data ($x \in \mathbb{R}^2$, $\mathrm{y} \in \mathbb{R}$). Hint: what are the support vectors?

1. $(0, 1)$
2. $(1,0)$
3. $(\frac{1}{2}, \frac{1}{2})$
4. $(\frac{6}{13}, \frac{4}{13})$

$x^{(1)}$	$x^{(2)}$	y
3	3	1
4	4	1
1	1	-1

The support vectors are the first and third data points. The SVM solution should lie between them and has the largest margin. Thus we have $w = \left(\frac{1}{2}, \frac{1}{2}\right)$.